# Fairness And Bias in Artificial Intelligence: A Brief Survey of Sources, Impacts, And Mitigation Strategies

Year: 2023 | Citations: 475 | Authors: Emilio Ferrara

## Abstract

The significant advancements in applying artificial intelligence (AI) to healthcare decision-making, medical diagnosis, and other domains have simultaneously raised concerns about the fairness and bias of AI systems. This is particularly critical in areas like healthcare, employment, criminal justice, credit scoring, and increasingly, in generative AI models (GenAI) that produce synthetic media. Such systems can lead to unfair outcomes and perpetuate existing inequalities, including generative biases that affect the representation of individuals in synthetic data. This survey study offers a succinct, comprehensive overview of fairness and bias in AI, addressing their sources, impacts, and mitigation strategies. We review sources of bias, such as data, algorithm, and human decision biases—highlighting the emergent issue of generative AI bias, where models may reproduce and amplify societal stereotypes. We assess the societal impact of biased AI systems, focusing on perpetuating inequalities and reinforcing harmful stereotypes, especially as generative AI becomes more prevalent in creating content that influences public perception. We explore various proposed mitigation strategies, discuss the ethical considerations of their implementation, and emphasize the need for interdisciplinary collaboration to ensure effectiveness. Through a systematic literature review spanning multiple academic disciplines, we present definitions of AI bias and its different types, including a detailed look at generative AI bias. We discuss the negative impacts of AI bias on individuals and society and provide an overview of current approaches to mitigate AI bias, including data pre-processing, model selection, and post-processing. We emphasize the unique challenges presented by generative AI models and the importance of strategies specifically tailored to address these. Addressing bias in AI requires a holistic approach involving diverse and representative datasets, enhanced transparency and accountability in AI systems, and the exploration of alternative AI paradigms that prioritize fairness and ethical considerations. This survey contributes to the ongoing discussion on developing fair and unbiased AI systems by providing an overview of the sources, impacts, and mitigation strategies related to AI bias, with a particular focus on the emerging field of generative AI.